MULTIVARIJATNE METODE

Mirjana Kujundžić Tiljak i Davor Ivanković

U prirodnim situacijama postoji mnoštvo varijabli koje su međusobno povezane. Govorimo o *multidimenzionalnosti pojava*, tj. o pojavama (događajima ili stanjima) koje su opisane s velikim brojem varijabli. Pritom je svaka varijabla posebna manifestna (mjerljiva) dimenzija. Primjereno je poželjeti te varijable (zbog dinamičkog ekvilibrija u kojem se nalaze) analizirati sve zajedno (kao entitet u cjelini) i istovremeno, a ne jednu po jednu ili par po par jer na taj način varijable bivaju izvučene iz njihovog realnog konteksta. Manje je, dakle, prirodno razbijati cjelinu na dijelove i promatrati ih neovisno od drugih varijabli. Analitičke postupke kojima analiziramo više varijabli istovremeno nazivamo *multivarijatnim ili multidimenzionalnim analitičkim postupcima.*

Ovaj pristup znači da se pažnja usmjerava prema $p(p-1)/2$ kovarijanci i varijanci u nekom skupu (p=broj varijabli) , a ne samo p aritmetičkih sredina i isto toliko varijanci. Npr. ako analiziramo 7 varijabli istraživač mora uzeti u obzir u svojoj analizi $7(7-1)/2 = 21$ veza među varijablama, a ako pak analiziramo 10 varijabli onda treba u analizi uzeti u obzir $10(10-1)/2 = 45$ veza među varijablama. Potom ih valja i interpretirati.

Neke od tih veza su bitne, neke nebitne, a neke su obični artefakti. Zbog složenosti veza među varijablama univarijatni statistički postupci često su ograničenog dometa. Često su čak i izvor zabluda, pogotovo ako se primjenjuju kada priroda istraživanog problema nalaže primjenu multivarijatnih analitičkih postupaka.

U analizi podataka prirodnih situacija dominiraju dva pitanja. Jedno je razumijevanje pojave koju promatramo, a drugo objašnjenje te pojave. Razumijevanje je u suštini odgovor na pitanje *ŠTO* (što se događa), a objašnjenje daje odgovor na pitanje *ZAŠTO* (zašto se to događa). Razumijevanje je pretpostavka svakog objašnjenja.

Konačno, prirodna istraživanja podrazumijevaju prostor i vrijeme. Prostor je multidimenzionalan, s često velikim brojem dimenzija (varijabli). Neki od multivarijatnih algoritama reduciraju prostor, od mnoštva varijabli među kojima postoji povezanost oni proizvode manji broj novih varijabli (tzv. latentnih varijabli ili dimenzija) pomoću kojih se uspješnije opisuje sustav nego li s originalnim, mjerenim, manifestnim varijablama. Drugi pristup koji uključuje komponentu vremena analizira promjene u sustavu (kvantitativne i kvalitativne), one koje se dešavaju spontano i one koje su izazvane izvjesnim intervencijama.

U literaturi se može naći obilje modela za analizu prirodnih situacija. To su u pravilu *multivarijatne analize podataka*. Multivarijatno mogu se analizirati sve vrste podataka, mjerenih na različitim skalama mjerenja.

Metode multivarijatne analize možemo kategorizirati na sljedeći način:

S obzirom na tipologiju u medicinskim istraživanjima u tekstu će se koristiti ova kategorizacija:

1. metode koje pretpostavljaju postojanje *jednog skupa varijabli* a odnose se na probleme proučavanja *strukture* tog skupa varijabli odnosno *entiteta* koji su opisani tim varijablama:
   - *FAKTORSKI MODEL*
   - *TAKSONOMSKI MODEL*

2. metode koje pretpostavljaju postojanje *najmanje dvaju skupova varijabli* a odnose se na probleme proučavanja *povezanosti i razlika* među tim skupovima varijabli:
   - *REGRESIJSKI MODEL*
   - *KANONIČKI MODEL*
   - *DISKRIMINACIJSKI MODEL*
   - *MANOVA (multivarijatna analiza varijance)*

3. metode za *analizu promjena* nekog skupa varijabli ili entiteta
   - VREMENSKE SERIJE

REGRESIJSKI MODEL

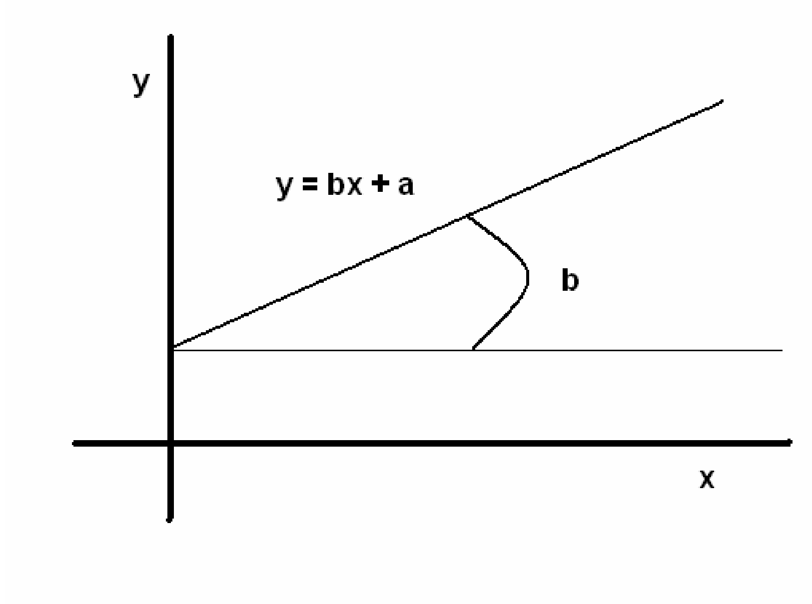
U medicinskim istraživanjima često se interes istraživača usmjerava prema problemu *povezanosti* među varijablama. Pri tom od posebnog interesa mogućnost *prognoziranja ili predikcije* vrijednosti (ili varijabilnosti) jedne varijable na osnovu drugih varijabli.

Prvi tako formulirani problem potječe od engleskog antropologa Francisa Galtona. On je studirajući zajedno sa Pearsonom nasljeđivanje u biologiji, mjereći visine očeva i sinova ustanovio neku vrstu paradoksa, odnosno, da visoki očevi imaju visoke sinove ali u prosjeku ne tako visoke kao što su oni sami, i slično, da niski očevi imaju niske sinove ali opet u prosjeku ne tako niske kao što su oni. Ovu tendenciju prosjeka neke karakteristike (u ovom slučaju visine) selekcionirane grupe da u sljedećoj generaciji sinova teži u prosjeku populacije a ne prosjeku njihovih očeva Galton je nazvao *regresijom*, točnije, *regresijom prema prosjeku*.

Da bi dobio informaciju o ovisnosti visine sinova od visine njihovih očeva Pearson je pretpostavio da se ta ovisnost može izraziti kao funkcija, $y = f(x)$, pri čemu je ***y*** *zavisna ili kriterijska varijabla* (u Galtonovom primjeru visina sinova) a *x nezavisna ili prediktorska varijabla* (visina očeva).

Ukoliko se radi o problemu zavisnosti koji se realno može opisati dvjema varijablama, jednom perdiktorskom (x) i jednom kriterijskom (y), zavisnost varijable y o varijabli x može se prikazati linearnom funkcijom od x:

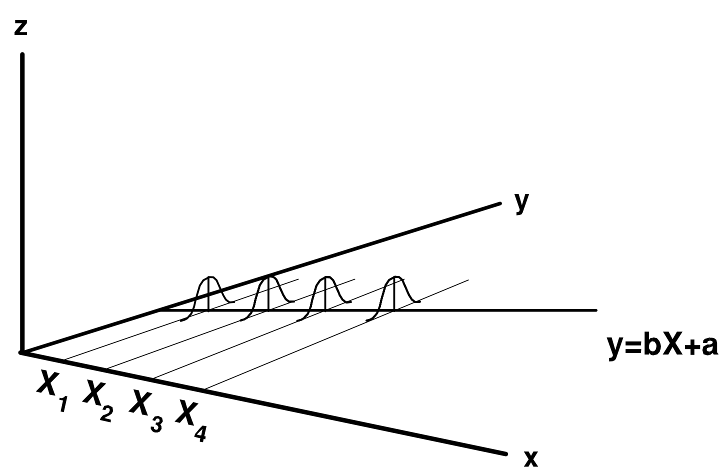
$$y = bx + a$$



Slika 1. Linearna regresija između dvije varijable

Parametar *b* je tangens kuta kojega pravac zatvara s pozitivnim smislom osi x ili nagib pravca. Nazivamo ga *koeficijentom regresije*, dok je *a* odsječak na ordinati. Jačina povezanosti ne zavisi od a, pa zato ukoliko ne postoje neki posebni razlozi, pravac se pomiče u ishodište Kartezijevog sustava (koordinatni sustav razapet apscisom i ordinatom) tako da je a = 0, a regresija prima oblik y = bx.

U realnim situacijama za istu vrijednost varijable x možemo imati više vrijednosti varijable y za koje se u regresijskoj analizi pretpostavlja da su normalno distribuirane. Može se pokazati da regresija tada prolazi prosjecima varijable y za svaku pojedinu vrijednost varijable x , što prikazuje slika 2.



Slika 2. Linearna regresija - multivarijatno

Opisani model ima primjenu u sustavima koje možemo opisati dvjema varijablama i gdje ne postoje utjecaji drugih nezavisnih varijabli x_2, x_3, ..., x_k, (one možda ne postoje ili ih možemo zanemariti). Kao primjer takvog sustava može se uzeti baždarni graf gdje se na osnovu vrijednosti x-a očitava vrijednost y-a, te u eksperimentima koji se razumno mogu opisati dvjema varijablama. Istraživanja prirodnih sustava su, međutim, složenija. Moramo pretpostaviti više nezavisnih varijabli tako da govorimo o prediktorskom sustavu, a zavisnost među varijablama prikazujemo *generalnim linearnim modelom*.

MULTIVARIJATNA REGRESIJA

Zbog složene prirodna istraživanja postoji *prediktorski skup varijabli*, tj. više nezavisnih varijabli $\mathbf{x_1, x_2, x_3, \ldots. x_n}$ te jedna *kriterijska ili zavisna varijabla* **y.** Zavisnost među varijablama prikazujemo *generalnim linearnim modelom.*
Varijabla y ne ovisi samo o varijablama $x_1 \ldots x_k$ nego i o nekim varijablama koje nam nisu dostupne. Dakle prognozirana ili procijenjena vrijednost y povezana je i s greškom:

$$y = f(x_1, x_2, x_3, \ldots. x_n) + e$$

Zbog pretpostavke linearnosti u parametrima ovaj model se naziva linearni model. Uobičajena oznaka za linearni model je: $(y, x, \beta, \sigma^2, I)$.

Objašnjenje:
y = vektor zavisnih ili kriterijskih varijabli
x = matrica dizajna, matrica vrijednosti za koje se pretpostavlja da su u ovisnosti od y, tzv. nezavisne ili prediktorske varijable
β= vektor nepoznatih parametara modela ili koeficijenata regresije
σ^2 = vektor greške > 0
I = jedinična matrica n-tog reda.

STANDARDNI REGRESIJSKI MODEL

Regresijski model prikazuje se na način:

$$w=Z\beta+e$$

Objašnjenje:
Z = matrica standardiziranih prediktora
w = vektor standardizirane kriterijske varijable

Opisani model svodi se na problem rješavanja sustava linearnih jednadžbi u kojima nam je parametar β je nepoznat. U tom modelu postoji serija *interkorelacija* između prediktorskih varijabli te prediktorskih sa kriterijskom varijablom. Tako dobijemo matricu korelacija R, čiji su elementi *Pearsonovi koeficijenti korelacija* koji nas informiraju o stupnju povezanosti između parova varijabli, a mogu primiti vrijednost od -1 do +1. Predznak govori da li se radi o pozitivnoj korelaciji (rastom vrijednosti jedne varijable, raste i vrijednost druge) odnosno negativnoj korelaciji (kada porastom vrijednosti jedne varijable, vrijednost druge opada). Ako su dvije varijable funkcionalno povezane (npr. opseg kruga $O = 2r\pi$) onda koeficijent korelacije prima vrijednost 1. Područje između -1 i +1 jest područje korelacija koje susrećemo u prirodnim i društvenim istraživanjima. Ta područja ne pretpostavljaju *determinirane* sustave sa funkcionalnim vezama među varijablama, nego *stohastičke* (vjerojatne). Matrica Pearsonovih koeficijenti korelacija u multivarijatnim analizama sirovina je (ili građevni materijal) na kojem se gradi neki analitički algoritam.

Kod primjene regresijske analize možemo naići na probleme. Naime, empirija je pokazala da regresijski koeficijenti, u slučaju malog uzorka ili čak i uzorka umjerene veličine (n reda veličine 100), znatno variraju od uzorka do uzorka istog osnovnog skupa. Zato uzorak treba biti dovoljno velik. Prilikom interpretacije regresijske analize ne treba gledati samo regresijske koeficijente, nego treba izračunati i *varijancu procjene kriterijske varijable* koja se obično označava kao R_d^2 i naziva *koeficijentom determinacije.*

Koeficijent determinacije R_d^2 može se interpretirati kao dio (proporcija) varijance koji se može objasniti promatranim prediktorskim sistemom. Može ga se smatrati mjerom efikasnosti regresije, odnosno uspješnosti prognoze. Signifikantnost koeficijenta determinacije R_d^2 (ili koeficijenta multiple korelacije R_d) testira se Fisherovim testom. Vrijednost R_d^2 varira od 0 do 1 odnosno, ako ga izražavamo kao procent, od 0 do 100%.

Daljnji važan dio regresijskog algoritma jest *multipla parcijalna korelacija*. Entiteti su, naime, opisani s mnoštvom varijabli koje su u međusobnoj korelaciji. Pri tome često korelacija (visoki koeficijenti korelacija) između dviju varijabli može biti posljedica djelovanja neke treće varijable koja je u visokoj korelaciji sa svakom od onih dviju. Zato se postavlja pitanje: kolika je stvarna korelacija između dviju varijabli, odnosno, možemo li na neki način iz izabranih dviju varijabli ukloniti utjecaj svih preostalih varijabli, tj. možemo li ukloniti „smetnju“ kako bismo dobili stvarnu povezanost izabranih varijabli? Odgovor je izračun multiple parcijalne korelacije.

Osim klasičnog linearnog modela regresijske analize postoje i drugi koji se javljaju u praksi analize podataka u medicinskim istraživanjima. Jedan tip modela dozvoljava i druge (nelinearne) odnose, a drugi dozvoljava da prediktorska i kriterijska varijabla ne budu numeričke prirode tj. omjernog ili intervalnog tipa.

Općenito se mogu promatrati najrazličitije ovisnosti, kao npr. eksponencijalne, polinomne i slično. Regresije s nelinearnošću u varijablama češće nalazimo kao rješenja u eksperimentalnim medicinskim istraživanjima ili onda kad postoji neko logičko opravdanje promatrati npr. x^2 ili x_1x_2 itd. kao varijable prediktore. Međutim, linearni model kao najjednostavniji apsolutno dominira u istraživanjima i praktički zadovoljava potrebe većine prirodnih i društvenih istraživanja.

Što se tiče problema varijabli koje nisu intervalnog ili omjernog tipa, ali se podaci mogu prikazati kao ordinalne varijable, te ukoliko se realno može pretpostaviti da je ispitivana varijabla normalno distribuirana, dopuštene su operacije *transformiranja originalnih vrijednosti*. Cilj je takve transformacije da ordinalne varijable prime svojstvo normalno distribuiranih, te s njima budu dopuštene sve statističke analize koje su primjerene originalno normalno distribuiranim varijablama. Takva se transformacija postiže tzv. normalizacijom varijabli pomoću i*nverznog normalnog integrala* čime varijabla postaje numerička i njezin prvobitni kvalitativni (nenumerički) karakter se gubi.

Daljnji problem u analizi podataka u medicinskim istraživanjima, kada je kriterijska odnosno zavisna varijabla binarna (y je 1 ili 0, infarkt: da ili ne), rješava se *logističkim regresijskim modelom*.

U analizama preživljavanja npr. bolesnika sa neoplazmama, koristi se *Coxov regresijski model*. Kriterijska varijabla je smrt u određenom vremenskom intervalu opisana sa funkcijom d (t).

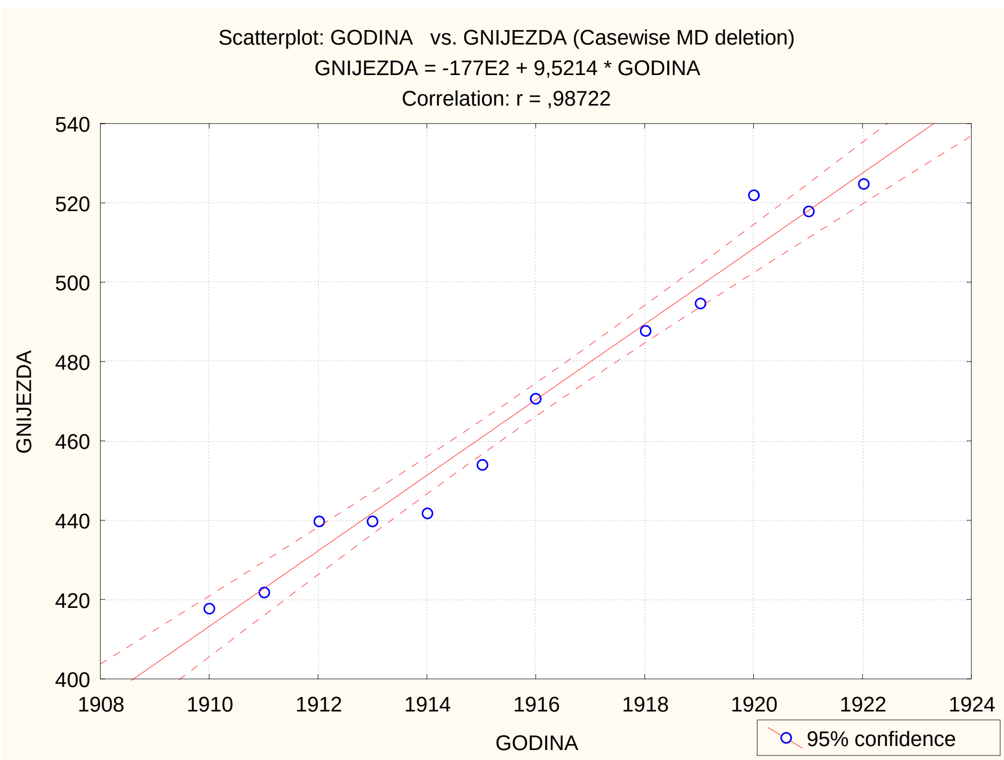
Poseban problem predstavljaju varijable koje nisu normalno distribuirane. Ukoliko su varijable bimodalno ili jako asimetrično distribuirane, regresijska analiza se ne bi smjela provoditi. Moguće rješenje je u pokušaju da se varijable transformacijama „normaliziraju“, bilo logaritamskom ili nekom drugom transformacijom.

Za regresijsku analizu poželjno je imati veće reprezentativne uzorke. U literaturi susrećemo i vrlo stroge zahtjeve kao na primjer da uzorak mora biti veći od 200. Ako je uzorak malen (<50) koeficijenti regresije mogu pokazivati izrazite slučajne fluktuacije. Prije regresijske analize treba osmišljeno izabrati prediktorski sustav koji ne smije sadržavati prediktorske varijable za koje je poznato da su u međusobnoj korelaciji. . Također treba obratiti pažnju na rezidualna odstupanja, tj. ponašanje greške e. Reziduumi moraju slijediti normalnu raspodjelu i ne smiju imati visoke vrijednosti.

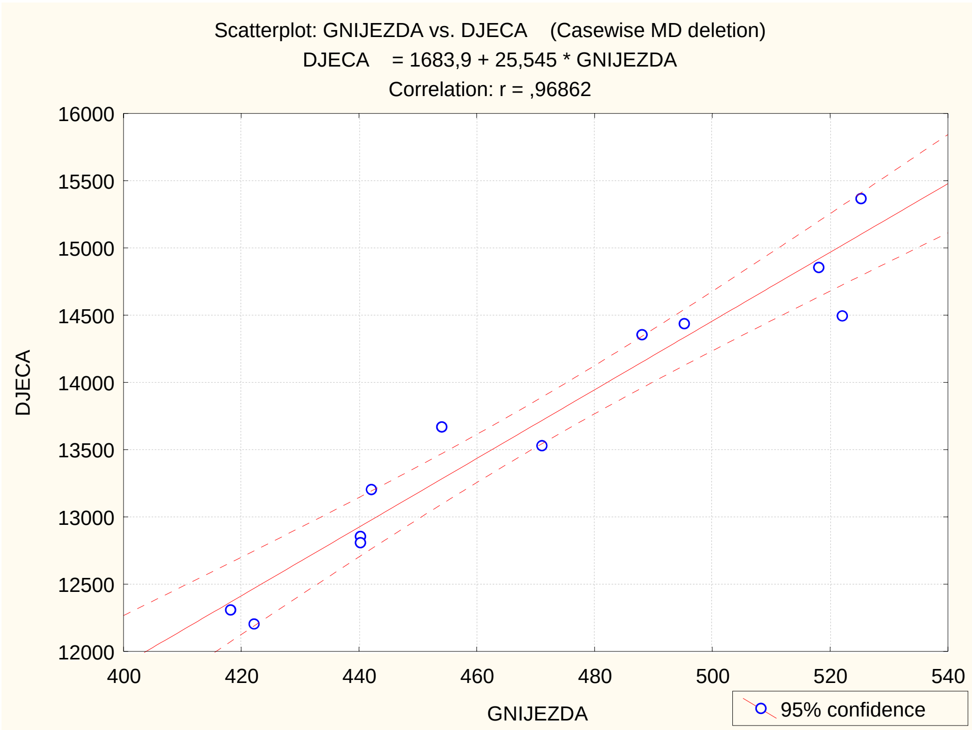
PRIMJER REGRESIJSKE ANALIZE

Tablica 1 Broj rodinih gnijezda i broj živorođene djece u kineskoj pokrajini Yukatan od 1920 do 1922 godine.

GODINA	GNIJEZDA	DJECA
1910	418	12312
1911	422	12208
1912	440	12857
1913	440	12819
1914	442	13204
1915	454	13670
1916	471	13538
1917		
1918	488	14364
1919	495	14437
1920	522	14503
1921	518	14866
1922	525	15376



Slika 3. Korelacija godina i rodinih gnijezda



Slika 4. Korelacija rodinih gnijezda i živorođene djece

Tablica 2: Rezultati regresijske analize

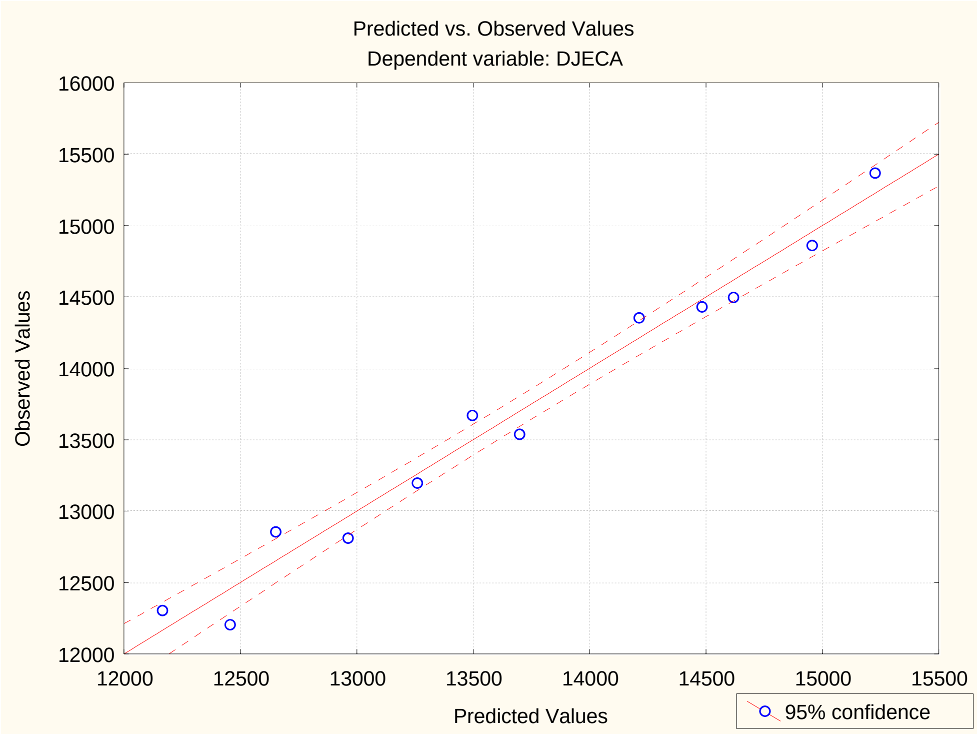
```
 Multiple Regression Results

Dependent: DJECA              Multiple R =  ,98825775     F = 188,2475
                                     R2=  ,97665339    df =   2,9
No. of cases: 12              adjusted R2=  ,97146525     p =  ,000000
             Standard error of estimate:174,25100455
Intercept: -582755,9078  Std.Error: 151828,2  t(    9) = -3,838  p =  ,0040

       GODINA beta=1,23       GNIJEZDA beta=-,25
```

Partial Correlations:

GODINA*DJECA:	0,788753	t(9)= 3,85; p=0,004
GNIJEZDA* DJECA:	-0,248351	t(9)=-0,78; p=0,462



Slika 5. Opažene i procijenjene vrijednosti varijable „živorođena djeca“

FAKTORSKA ANALIZA

Faktorska analiza je generičko ime za mnoštvo procedura razvijenih sa svrhom analize interkorelacija unutar jednog skupa varijabli i posljedične redukcije dimenzionalnosti prostora. Jedan od glavnih razloga za korištenje faktorske analize u znanosti je tzv. *zakon štednje* (engl. law of parsimony). Zahtjev za parsimonijskim rješenjem problema predstavlja zahtjev da se što veći broj varijabli objasni pomoću što manjeg broja varijabli. Stoga je i cilj faktorske analize da međusobnu povezanost većeg broja varijabli objasni nekim manjim brojem fundamentalnih ili latentnih varijabli, odnosno dimenzija, odnosno izvora kovarijacija.

Dva su tipa faktorskih procedura: eksploratorna faktorska analiza i konfirmatorna faktorska analiza. *Eksploratorna faktorska analiza* trebala bi omogućiti identifikaciju onoga što u podacima o modelu objektivno postoji. Ona omogućuje da se utvrde temeljni faktori odnosno izvori varijacija i kovarijacija među promatranim (manifestnim) varijablama. *Konfirmatorna faktorska analiza* pretpostavlja postojanje unaprijed formuliranog modela, hipoteze ili teorije o strukturi temeljnih izvora varijacija i kovarijacija među promatranim varijablama. Ta se hipoteza izražava u obliku *matrice cilja, ciljne matrice* (engl. target matrix), a zatim se provodi testiranje da li se empirijski podaci slažu s hipotetičkim.

Faktorska analiza raspoloživi prostor originalnih *manifestnih varijabli* aproksimira prostorom koji mu je dovoljno blizak, nosi otprilike *istu količinu informacija* kao originalni prostor ali manji broj dimenzija, manji broj varijabli koje ga opisuju. Pri redukciji dimenzionalnosti prostora osnovno je da se iz *p dimenzija* (originalnih, mjerenih, manifestnih) originalnog prostora ekstrahira *k ($k \leq p$) linearnih kombinacija* tih varijabli *(latentnih varijabli)* koje će u signifikantnoj proporciji objasniti *ukupnu varijancu.*

Zadaci faktorske anlize su:

a) utvrditi *faktore* koji leže u osnovi međusobne povezanosti manifestnih varijabli, tj. reducirati dimenzionalnost originalnog prostora, nekim postupkom faktorizacije;

b) utvrditi *povezanost pojedinih manifestnih varijabli s tim faktorima* tj. *rotacijom* faktora postići interpretabilnija rješenja.

METODA GLAVNIH KOMPONENTI (engl. Principal component analysis)

Analiza glavnih komponenata je prva i klasična metoda kojom pokušavamo reducirati dimenzionalnost tako da konstruiramo latentne varijable koje su međusobno nezavisne(nisu međusobno u korelaciji) a onda zadržimo samo one koje su "dovoljno informativne". Razvio ju je Hotelling, 1933. godine.

Da bi se reducirao broj varijabli odnosno dimenzija prostora, potrebno je pronaći kriterij za odbacivanje "malo informativnih" novih varijabli odnosno za zadržavanje onih varijabli koje nose najveći dio informacija sadržanih u polaznom sustavu varijabli. Postoji niz kriterija a neki od njih su:

1. Kaiserov kriterij prema kojemu se uzimaju samo one glavne komponente kojima odgovaraju svojstvene vrijednosti veće od 1;

2. uzimaju se samo one glavne komponente kojima odgovaraju svojstvene vrijednosti veće od prosjeka svih svojstvenih vrijednosti;

3. uzimaju se samo one glavne komponente koje nose odnosno zadržavaju svaka posebno unaprijed zadani dio informacija;

4. uzimaju se samo one glavne komponente koje nose odnosno zadržavaju (ukupno kao sustav) unaprijed zadani dio informacija;

5. broj komponenata se određuje proizvoljno.

Nema pravila koje bi diktiralo izbor jednog od navedenih kriterija. Jedino čime se treba rukovoditi jest cilj da broj zadržanih varijabli ne bude prevelik i da gubitak informacija iz polaznog sustava bude što manji.

U interpretaciji novodobivenih varijabli služimo se *matricom strukture* čiji su elementi linearne korelacije originalnih varijabli i novodobivenih varijabli ili *faktora*.

ROTACIJA FAKTORA

U praksi multivarijatne analize podataka rijetko se istraživači zadržavaju samo na razini nalaženja glavnih komponenata. Razlog leži u činjenici da se "klasteri" varijabli ne mogu uvijek lagano prepoznati primjenom samo modela glavnih komponenata. Faktori često nisu definirani tako da se jedna varijabla javlja samo na jednom faktoru. Naprotiv, događa se da se *jedna varijabla javlja na više faktora*, tako da imamo varijable koje definiraju veći broj faktora. Da bi se to izbjeglo nastoji se dobivene glavne komponente *transformirati (rotacijom pod određenim uvjetima)* da bi se dobila interpretabilnija rješenja. Cilj takvih transformacija jest dobivanje *jednostavne strukture* koja potječe od Thurstonea (1947) znači, da faktori trebaju biti što nezavisniji, tj. jedan bi faktor trebao biti određen(ili opisan) jednim skupom varijabli, drugi drugim itd., i da pritom bude što manje varijabli koje bi bile zajedničke većem broju faktora.

Jedan kriterij za transformaciju faktora koji osigurava dobivanje jednostavne strukture za faktorsku matricu je *Kaiserov varimax kriterij*. Ovaj kriterij osigurava simplifikaciju stupaca. To znači da za svaki faktor *varimax (ortogonalna) rotacija* teži dati veliko opterećenje (velike vrijednosti elemenata u faktorskoj matrici, velike korelacije između faktora i varijabla) malom broju manifestnih varijabli. Ostatak opterećenja teži da bude što bliže nuli.

Za razliku od varimax rotacija i drugih *ortogonalnih (pravokutnih) rotacija* postoje i *kose rotacije* koje napuštaju zahtjev da faktorske osi moraju biti pod pravim kutom tj. ortogonalne. Razlog za tu vrstu rotacija leži u činjenici da se ponekad "klasteri" varijabli ne mogu prepoznati u ortogonalnoj poziciji.

Kako se radi o kosoj projekciji sada imamo dvije vrste projekcije, jednu paralelnu s faktorskim osima (x_1 i x_2) i drugu okomitu na faktorske osi (w_1 i w_2). Prvu projekciju zovemo projekcijom strukture, a drugu projekcijom sklopa. Te projekcije dobivamo u rezultatima faktorske analize kao *matrice strukture (engl. structure matrix)* i *matrice sklopa (engl. pattern matrix)*. Elementi matrice strukture su *koeficijenti korelacija pojedinih manifestnih varijabli s faktorima*. Elementi matrice sklopa su tzv. *faktorska opterećenja ili faktorski ponderi,* za koje se može reći da odgovaraju koeficijentima regresije pojedine manifestne varijable na svaki faktor.
Jasno je da elementi matrice sklopa mogu biti veći od 1, dok se elementi matrice strukture mogu se kretati između -1 i +1.

U interpretaciji faktorske analize navodi se i koliko je *ukupno varijance* u sustavu od ***p*** varijabli objašnjeno s ***k*** zadržanih komponenta, dok ***komunalitet*** pojedine varijable govori koliko je varijance određene varijable objašnjeno s ***k*** zadržanih komponenata (faktora)

PRIMJER FAKTORSKE ANALIZE

“OPIS OBITELJSKOG LIJEČNIKA“

- multidimenzionalni prostor od 31 osnovne manifestne varijable reduciran je na 10 latentnih dimenzija, međusobno nezavisnih varijabli, glavnih komponenata, faktora sa svojstvenim vrijednostima većim od 1
- u sustavu je zadržano 66,8% informacija koje primarno nosi početni sustav od 31 manifestne varijable
- primijenjena je Varimax rotacija (ortogonalna)

Tablica 1. Komunaliteti manifestnih varijabli

31 manifestna varijabla:	KOMUNALITETI:
• 'udaljenost do liječnika opće medicine'	0,549
• 'udaljenost do specijaliste konzultanta'	0,645
• 'udaljenost do bolnice'	0,712
• 'sastav radne jedinice'	0,754
• 'dogovaranje pregleda'	0,638
• 'dogovaranje s liječnikom opće medicine'	0,629
• 'dogovaranje sa specijalistom konzultantom'	0,675
• 'dogovaranje s bolničkim specijalistom'	0,686
• 'dogovaranje s ljekarnikom'	0,439
• 'dogovaranje s patronažnom sestrom'	0,480
• 'dogovaranje s ambulantnom sestrom'	0,748
• 'dogovaranje sa socijalnim radnikom'	0,406
• 'dob'	0,902
• 'dužina specijalističkog staža'	0,776
• 'veličina mjesta'	0,651
• 'broj pacijenata na listi'	0,746
• 'broj liječnika u radnoj jedinici"	0,745
• 'broj kontakta s pacijentom u ambulanti'	0,743
• 'broj kontakta s pacijentom u kućnoj posjeti'	0,425
• 'broj kontakta s pacijentom telefonom'	0,636
• 'vrijeme čekanja na pregled'	0,538
• 'udio predškolske djece na listi pacijenta'	0,558
• 'udio ostarjelih na listi pacijenta'	0,641
• 'udio socijalno ugroženih na listi pacijenta'	0,651
• 'udio doseljenih na listi pacijenta'	0,709
• 'trajno usavršavanje - obnova znanja'	0,498
• 'ukupno radno iskustvo'	0,928
• 'radno iskustvo u općoj medicini'	0,837
• 'standardni paket opreme'	0,776
• 'napredni paket opreme'	0,747
• 'dodatni paket opreme'.	0,842

Tablica 2. Faktorska matrica

	FAKTOR 1	FAKTOR 2	FAKTOR 3	FAKTOR 4	FAKTOR 5
V1	-,05517	**,61976**	-,12517	-,00685	-,31404
V2	-,08215	**,67826**	,02314	,00272	-,34314
V3	-,06518	**,72181**	,21702	,21191	,06657
V4	,00181	-,25653	,17277	-,03432	**,79054**
V5	,03927	,03291	,01855	-,05785	,11520
V6	-,04907	-,07295	,12437	,17157	,14883
V7	,02363	,02726	,14488	**,77021**	,09326
V8	,09449	,05164	-,05188	**,81393**	-,00552
V9	,05560	,20039	-,03810	**,51933**	-,13454
V10	,12775	,17492	,00674	,17077	,01773
V11	,00878	-,01807	-,10924	,09455	-,09464
V12	,09885	-,06655	,14594	,48455	-,05754
V13	**,94023**	-,00487	-,04504	,08181	,05398
V14	**,86164**	-,07897	,01286	,06745	-,01764
V15	-,01815	**-,71660**	-,11060	-,09936	,11585
V16	,14273	-,01000	,04114	,06535	,08611
V17	,01917	-,30149	,12982	-,02124	**,78530**
V18	-,04974	,05674	-,03430	-,00729	-,01192
V19	,13389	,17485	,12190	-,01293	-,20522
V20	-,01885	-,37098	-,01001	,31203	-,31385
V21	,07335	,12254	,17463	,02591	,14568
V22	-,00865	**,55174**	,00520	-,07724	,04249
V23	,03583	,05010	,05189	-,08004	,06345
V24	,06538	,16895	,01194	,02432	,02662
V25	,07869	-,07836	,08305	-,01718	-,01462
V26	-,04998	-,18203	,27433	-,01809	-,49775
V27	**,95809**	-,02798	-,00984	,06251	,01074
V28	**,89602**	-,01292	,10173	,02713	,01157
V29	,06829	,20278	**,82232**	-,07036	,00930
V30	-,09911	-,01761	**,80907**	,13195	,06359
V31	,08696	-,00050	**,89999**	,08459	,09600

Tablica 2. Faktorska matrica (nastavak)

	FAKTOR 6	FAKTOR 7	FAKTOR 8	FAKTOR 9	FAKTOR 10
V1	,08312	-,10183	-,09841	-,13979	,02381
V2	-,13616	,18675	,04708	,03613	-,05271
V3	-,08774	,01781	,05661	,27848	,04067
V4	,06225	,16825	,01296	,00803	-,00976
V5	,09616	**,67104**	-,20765	,27676	-,19832
V6	**,72843**	,12129	,03046	-,03825	-,08385
V7	,17212	-,09144	-,05109	-,09409	,03348
V8	-,02155	-,03593	,00168	,04341	-,07674
V9	,20620	,16334	,13831	,12841	,03783
V10	**,46730**	,05753	,26775	,30056	,13976
V11	**,82897**	-,17269	-,01889	,00141	-,01349
V12	,12297	,12422	,03356	-,31296	,05502
V13	-,00254	-,00226	,02512	-,02987	,07249
V14	,05628	,07792	,08392	,06660	,03407
V15	-,11058	,12537	,01970	-,27030	-,00445
V16	-,05646	-,16006	**,79690**	,16617	-,14405
V17	-,07045	,10796	,01035	,04175	,03304
V18	,13685	,07856	**,83323**	-,03657	,12595
V19	,09884	,03231	,20201	**,51314**	,06755
V20	-,03676	**,54224**	,00236	-,00433	-,08371
V21	-,07609	**,65920**	,06351	-,05765	,13149
V22	-,02616	,37674	,24201	-,14794	,14899
V23	-,00266	,06832	-,00887	**,77135**	,15601
V24	-,08396	-,17937	-,05325	,16990	**,73862**
V25	,04216	,13429	,06243	,05399	**,81391**
V26	-,11472	,19965	-,08949	,27728	,03177
V27	,03373	,02102	,04024	-,02463	,03969
V28	-,06069	-,00301	-,04766	,12844	,01953
V29	,18546	,09240	,05549	-,03987	,03671
V30	-,18257	,02445	-,02386	,16110	,02680
V31	,02092	,06421	-,01645	,03322	,04574

Tablica 3.
Opis 10 latentnih varijabli (faktora) koji se mogu nazvati *"faktori osobe i prakse"*

- *F1- "radno iskustvo"*
  dob (+)
  dužina specijalističkog staža (+)
  ukupno radno iskustvo(+)
  radno iskustvo u općoj medicini (+)

- *F2 – "smještaj ambulante"*
  udaljenost do liječnika opće medicine (+)
  udaljenost do specijaliste konzultanta (+)
  udaljenost do bolnice (+)
  udio predškolske djece na listi pacijenata (+)
  veličina mjesta (-)

- *F3 – "oprema ambulante"*
  standardni paket opreme (+)
  napredni paket opreme (+)
  dodatni paket opreme (+)

- *F4 - "suradnja izvan tima"*
  dogovaranje sa specijalistom konzultantom (+)
  dogovaranje s bolničkim specijalistom (+)
  dogovaranje s ljekarnikom (+)

- *F5 - "grupiranje liječnika"*
  sastav radne jedinice (+)
  broj liječnika u radnoj jedinici (+)

- *F6 –"suradnja unutar tima"*
  dogovaranje s liječnikom opće medicine (+)
  dogovaranje s patronažnom sestrom (+)
  dogovaranje s ambulantnom sestrom (+)

- *F7 - "pristupačnost ambulante"*
  dogovaranje pregleda (+)
  broj kontakata s pacijentom telefonom (+)
  vrijeme čekanja na pregled (+)

- *F8 - "radno opterećenje ambulante"*
  broj pacijenata na listi (+)
  broj kontakata s pacijentom u ambulanti (+)

- *F9 - "skrb za ostarjele"*
  broj kontakata s pacijentom u kućnoj posjeti (+)
  udio ostarjelih na listi pacijenta (+)

- *F10 – "specifične zdravstvene potrebe"*
  udio socijalno ugroženih na listi pacijenta (+)
  udio doseljenih na listi pacijenta (+)

DISKRIMINACIJSKA ANALIZA

Diskriminacijska analiza vodi se idejom da ustanovi koje varijable prave najveću razliku među uspoređivanim grupama entiteta. Polazi se dakle od nekoliko grupa entiteta opisanih nizom varijabli. Zahtijeva se da se konstruiraju nove varijable (kojih treba biti manje nego polaznih) koje bi opisale razlike među grupama. Te nove varijable zovu se *diskriminacijskim varijablama (ili funkcijama)*. One se dobivaju kao linearne kombinacije izvornih varijabli, formiraju se prema zahtjevu da što bolje razlikuju grupe a interpretiraju se na temelju odnosa (korelacija) originalnih i diskriminacijskih varijabli.

Glavni zadaci diskriminacijske analize su prema tome slijedeći:
- *određivanje diskriminacijskih varijabli* tj. varijabli na kojima se grupe međusobno što je moguće više razlikuju;
- *redukcija broja diskriminacijskih varijabli* tj. zadržavanje samo onih varijabli na kojima se centroidi pojedinih grupa značajno razlikuju;
- *interpretacija diskriminacijskih varijabli pomoću originalnih*.

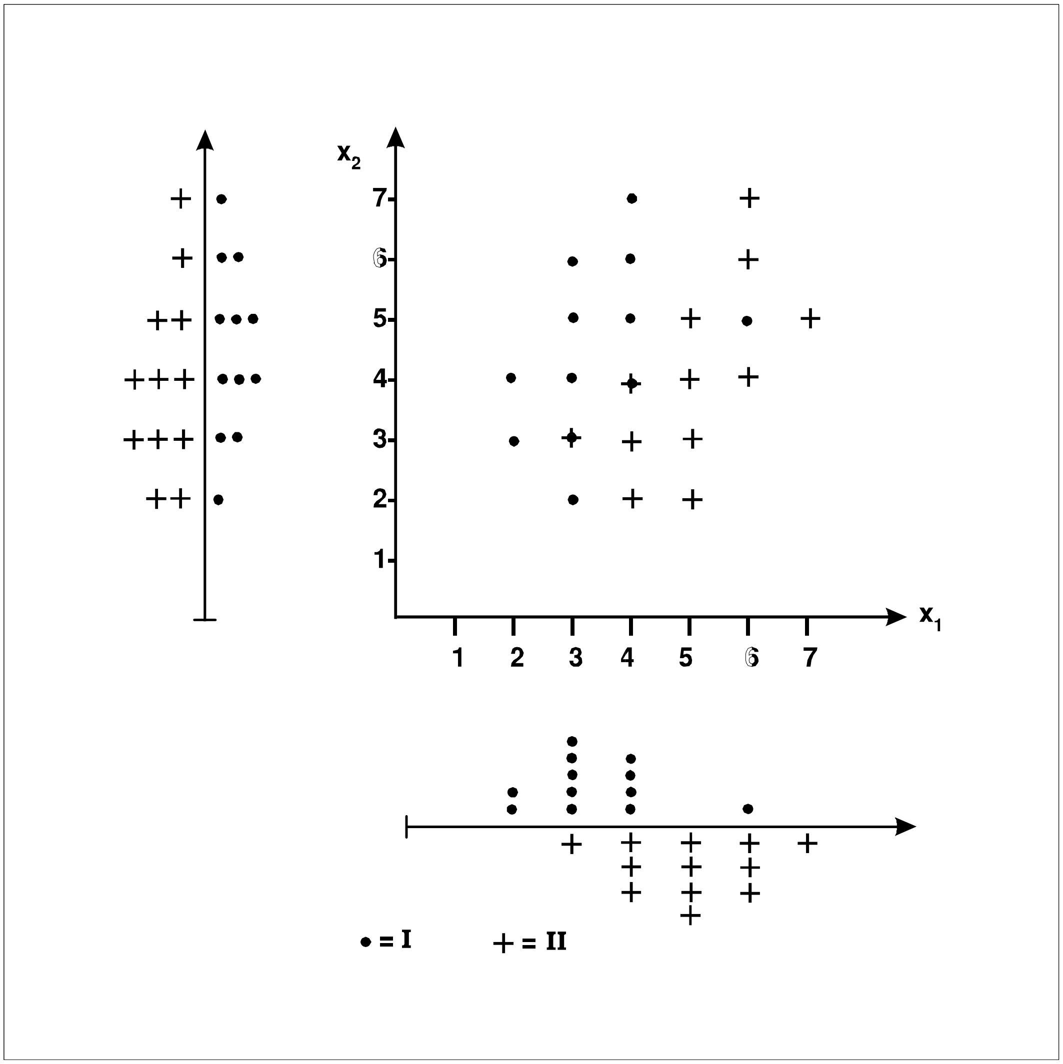
Postoji više različitih algoritama za diskriminacijsku analizu. Neki od tih algoritama na temelju diskriminacijskih varijabli prognoziraju pripadnost pojedinih entiteta pojedinim grupama. Zato i kažemo da originalne varijable (koje opisuju entitete i iz kojih se izvode diskriminacijske varijable) predstavljaju *prediktore* dok selektorski skup (binarnih varijabli) kojim određujemo pripadnost entiteta pojedinoj grupi zovemo *kriterijskim skupom*.

Slike 1 i 2 ilustriraju ideju diskriminacijske analize (u bivarijatnom slučaju, jer opći multivarijatni slučaj nije moguće grafički prikazati). Zadane su dvije varijable x_1 i x_2 i dvije grupe entiteta (I i II) opisane tim varijablama. Potrebno je na osnovi vrijednosti varijabli x_1 i x_2 odrediti pripadnost pojedinih entiteta pojedinim grupama.

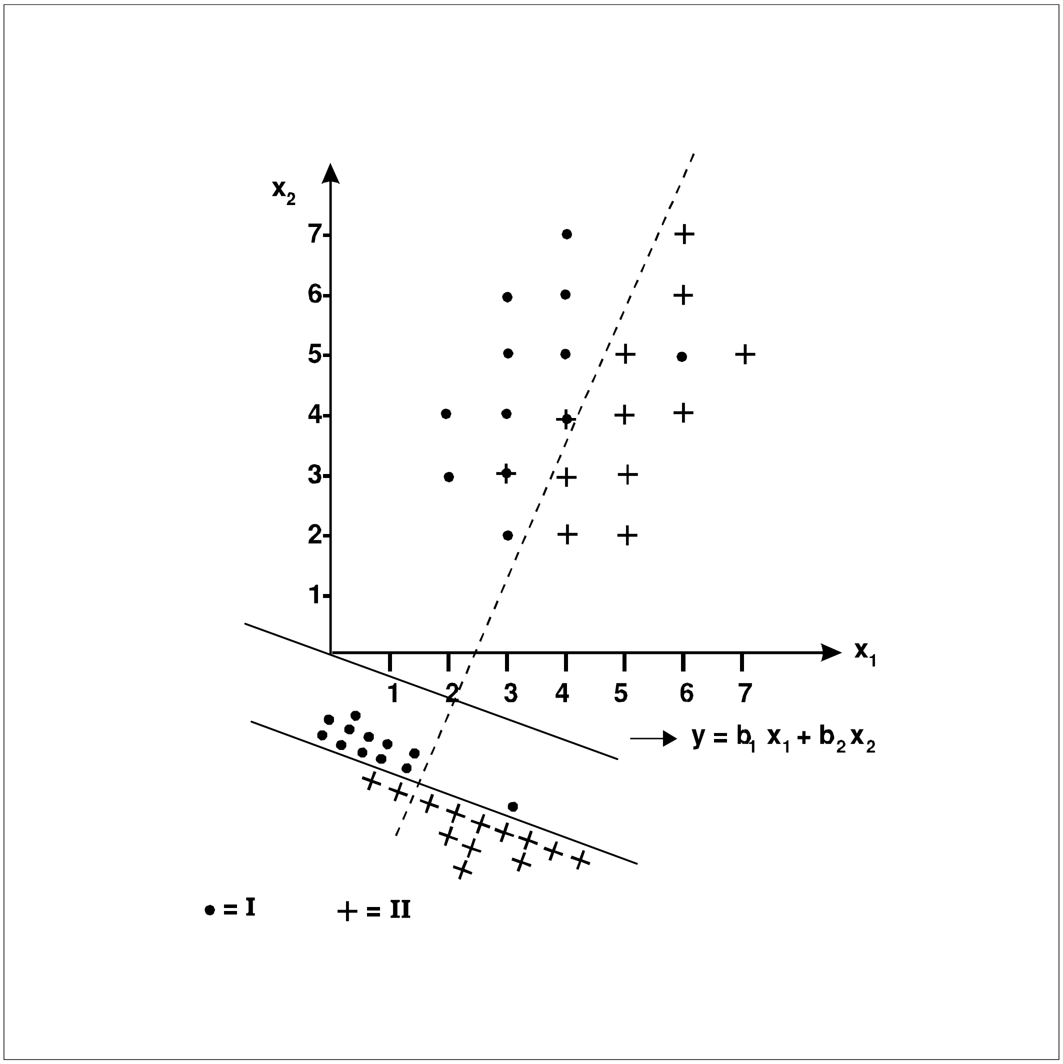
Grupe entiteta su prikazane pripadnim bivarijatnim distribucijama (varijabli) koje su projicirane na ravninu određenu sa x_1 i x_2.

Sa slike 1 se vidi da neki entitet (ispitanik) koji ima vrijednost varijable x_1 u intervalu (1,3) pripada grupi I. Ako je pak vrijednost varijable $x_1 = 7$ onda taj entitet pripada grupi II. Međutim, ako je vrijednost varijable x_1 iz intervala (4,6) onda na temelju samo te vrijednosti ne možemo reći da li je ispitanik iz grupe I ili II, nego treba promatrati i vrijednost varijable x_2. Ideja je diskriminacijske analize da se konstruira neka nova varijabla y (diskriminacijska varijabla) koja diskriminira grupe što je moguće bolje, tj. varijabla koja ima najmanji mogući interval preklapanja. Sa slike 2 se vidi da varijabla y ima tražena svojstva. Nadalje, broj varijabli (kojima opisujemo entitete) smo reducirali, pa umjesto dvije varijable x_1 i x_2 promatramo samo jednu varijablu y. Tu varijablu treba interpretirati, a to se radi na isti način kao i kod metoda opisanih u poglavlju Faktorska analiza. U interpretaciji se, naime, koriste relacije između originalnih varijabli x_1 i x_2 i diskriminacijske varijable y.

Slika 6



Slika 7



PRIMJER DISKRIMINACIJSKE ANALIZE

“DJELOKRUG OBITELJSKOG LIJEČNIKA“

203 ispitanika podijeljenih u 2 skupine:

1. skupina: 115 specijalista opće medicine
2. skupina: 88 liječnika opće medicine bez specijalizacije

14 originalnih varijabli:
- *varijable koje opisuju djelokrug rada liječnika obiteljske medicine:*

"primjena medicinske tehnike u vlastitoj praksi“
"prvi susret u slučaju zdravstvenih problema"
"liječenje i praćenje bolesti"
"preventiva i drugi postupci“
- *varijable koje opisuju osobu i praksu ("faktori osobe i prakse“):*

"radno iskustvo" (1. faktor),
"smještaj ambulante" (2. faktor)
"oprema ambulante" (3. faktor)
"suradnja izvan tima" (4. faktor)
"grupiranje liječnika" (5. faktor)
"suradnja unutar tima" (6. faktor)
"pristupačnost ambulante" (7. faktor)
"radno opterećenje ambulante" (8. faktor)
"skrb za ostarjele" (9. faktor)
"specifične zdravstvene potrebe" (10. faktor)

Tablica 4. Standardizirani koeficijenti kanoničkih diskriminacijskih funkcija

	funkcija 1
"primjena medicinske tehnike"	-,18979
"prvi susret"	,34928
"liječenje i praćenje bolesti"	,06466
"preventiva i drugi postupci"	,20899
"radno iskustvo" (F1)	,94424
"smještaj ambulante" (F2)	-,17744
"oprema ambulante" (F3)	-,05575
"suradnja izvan tima" (F4)	,15982
"grupiranje liječnika" (F5)	,18741
"suradnja unutar tima" (F6)	,13396
"pristupačnost ambulante" (F7)	,06509
"radno opterećenje ambulante" (F8)	,22768
"skrb za ostarjele" (F9)	,06775
"specifične zdrav. potrebe" (F10)	-,11288

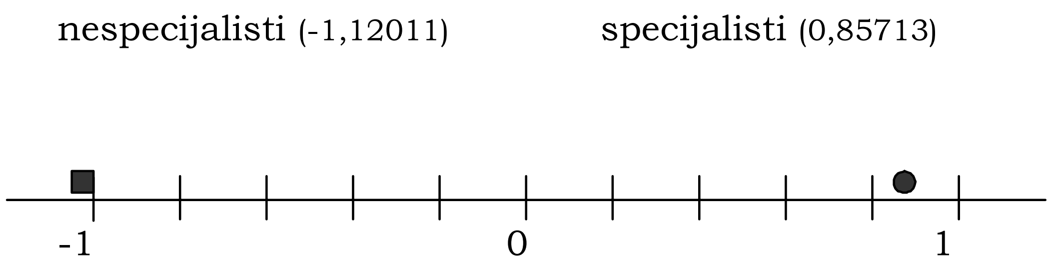
Tablica 5. Matrica strukture

	funkcija 1
"radno iskustvo" (F1)	***,83615***
"liječenje i praćenje bolesti"	***,26935***
"prvi susret"	***,22242***
"preventiva i drugi postupci"	***,17583***
"radno opterećenje ambulante" (F8)	***,12712***
"suradnja izvan tima" (F4)	,08536
"suradnja unutar tima" (F6)	,08045
"smještaj ambulante" (F2)	-,07735
"primjena medicinske tehnike"	,06732
"grupiranje liječnika" (F5)	,06324
"pristupačnost ambulante" (F7)	,05669
"specifične zdrav. potrebe" (F10)	-,04790
"skrb za ostarjele" (F9)	,04536

Rezultat diskriminacijske analize = diskriminacijska funkcija na koju se dobro projiciraju sljedeće varijable:

"radno iskustvo" (1. faktor) (+)
"liječenje i praćenje bolesti" (+)
"prvi susret" (+)
"preventiva i drugi postupci" (+)
"radno opterećenje ambulante" (+)

Slika 8. Centroidi grupa ispitanika udvodimenzionalnom diskriminacijskom prostoru



Zaključno, prema rezultatu diskriminacijske analize specijalisti opće medicine (za razliku od nespecijalista) su:

– stariji
– imaju duži specijalistički staž
– imaju veće ukupno radno iskustvo
– imaju veće radno iskustvo u općoj medicini
– više liječe i prate bolesti
– Imaju više prvih susreta u slučaju zdravstvenih problema
– više primjenjuju preventivu i druge postupke
– imaju veći broj pacijenta na listi
– veći broj kontakta s pacijentom u ambulanti.

Literatura:

1. *Ivanković D, i sur. Osnove statističke analize za medicinare. Zagreb: Medicinski fakultet Sveučilišta u Zagrebu, 1989.*
2. *Petrie A, Sabin C. Medical Statistics at a Glance (2nd Ed). Oxford: Blackwell Science Ltd, 2005.*
3. *Glantz. SA. Primer of Biostatistics (4th Ed). New York: McGraww-Hill: 1997.*
4. *Altman DG. Practical Statistics for Medical Research. London. Chapman & Hall, 1991.*
5. *Bland M. An Introduction to Medical Statistics (3rd Ed). Oxford: Oxford University Press, 2005.*
6. *Armitage P, Berry P. Statistical Methods in Medical Research. Oxford: Blackwell Science Ltd, 1994.*